ACE Scoring for 2005

- **Scorer** overview
- Pilot annotation scores

The Scoring Method

- The **scorer** scores the performance of a system by computing the "value" of the system's output using a three-step process:
  1. The value of each system output element is computed for all possible sys-ref mappings, including misses (sys doesn't match any ref) and false alarms (ref doesn't match any sys).
  2. An optimum association (one-to-one mapping) of sys elements to ref elements is found so that the resulting bottom-line score is maximized.
  3. The bottom-line score is computed, along with a myriad of diagnostic information, including various attribute-conditioned performance statistics and various attribute confusion matrices.

"Cross-document" = "Real World"

- The **scorer** scores the value of inferences made about things in the "real world".
  - To be valuable these inferences must represent real-world elements that exist separate and apart from the document(s) that mention them.
  - Real world identity is represented in the **apf** file by means of a globally unique ID that is assigned to each (unique) real-world element.
  - The **scorer** always performs real-world scoring (i.e., "cross-document" scoring).
  - If document-level scoring is desired, then every element must be mentioned in only one document. This may be achieved by severing cross-document links in the reference data (by assigning a unique element ID to each set of element mentions that are restricted to a single document).

The EDR Scoring Formula

$$EDR\_Value_{sys} = \sum_i value\_of\_sys\_entity_i$$

$$value\_of\_sys\_entity = Entity\_Value(sys\_entity) \cdot \sum_m Mention\_Value(sys\_men_m \Rightarrow ref\_men_{map(m)})$$

$$Entity\_Value = \begin{cases} \min\begin{pmatrix} EClassValue(sys), \\ EClassValue(ref_{sys}) \end{pmatrix} \cdot \prod_i W_{Eerr-attribute(i)} & \text{when mapped} \\ EClassValue(sys) \cdot W_{E-FA} & \text{when entity is not mapped} \end{cases}$$

$$Mention\_Value = \begin{cases} \min\begin{pmatrix} MTypeValue(sys\_men), \\ MTypeValue(ref\_men_{mapped}) \end{pmatrix} \cdot \prod_i W_{Merr-attribute(i)} & \text{when mapped} \\ -MTypeValue(sys\_men) \cdot (W_{M-FA} \cdot W_{M-CR}) & \text{when mention isn't mapped} \end{cases}$$

The QDR Scoring Formula

$$QDR\_Value_{sys} = \sum_i value\_of\_sys\_quantity_i$$

$$value\_of\_sys\_quantity = Quantity\_Value(sys\_quantity) \cdot \sum_m Mention\_Value(sys\_men_m \Rightarrow ref\_men_{map(m)})$$

$$Quantity\_Value = \begin{cases} \min\begin{pmatrix} QTypeValue(sys), \\ QTypeValue(ref_{sys}) \end{pmatrix} \cdot \prod_i W_{Qerr-attribute(i)} & \text{when mapped} \\ QTypeValue(sys) \cdot W_{Q\text{-}FA} & \text{when quantity not mapped} \end{cases}$$

$$Mention\_Value = \begin{cases} 1 & \text{when mapped} \\ -W_{Q-FA} & \text{when mention not mapped} \end{cases}$$

The RDR Scoring Formula

$$RDR\_Value_{sys} = \sum_i value\_of\_sys\_relation_i$$

$$value\_of\_sys\_relation = Relation\_Value(sys\_relation) \cdot \sum_a Argument\_Value(sys\_arg_a \Rightarrow ref\_arg_{map(a)})$$

$$Relation\_Value = \begin{cases} \prod_i W_{Rerr-attribute(i)} & \text{when mapped} \\ W_{R\text{-}FA} & \text{when relation not mapped} \end{cases}$$

$$Argument\_Value = Element\_Value(sys\_arg \Rightarrow ref\_arg) + \begin{pmatrix} Element\_Value(sys\_arg \Rightarrow ref\_arg) - \\ Element\_Value(sys\_arg \Rightarrow sys\_arg) \end{pmatrix} \cdot W_{A-FA}$$

The VDR Scoring Formula

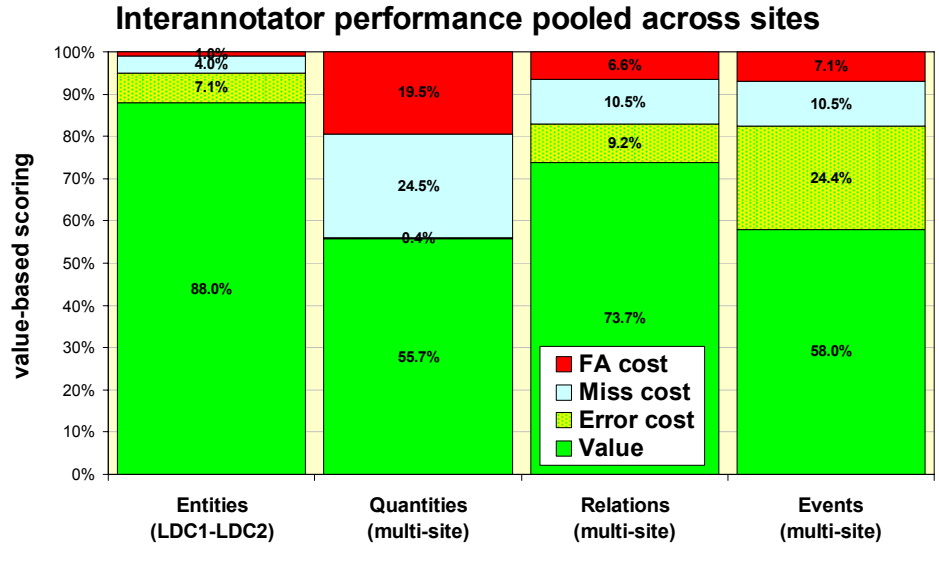
$$VDR\_Value_{sys} = \sum_i value\_of\_sys\_event_i$$

$$value\_of\_sys\_event = Event\_Value(sys\_event) \cdot \sum_a Argument\_Value(sys\_arg_a \Rightarrow sys\_arg_{map(a)})$$

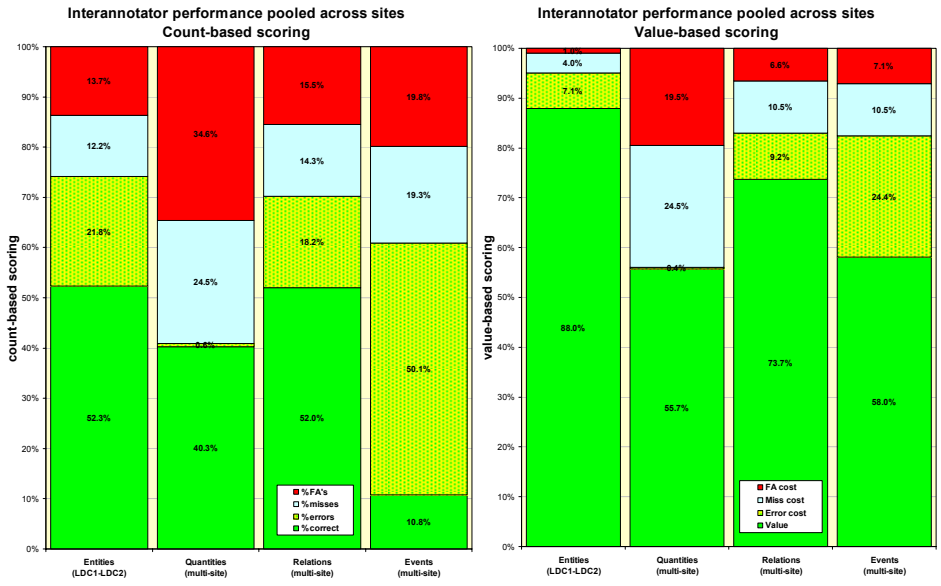
$$Event\_Value = \begin{cases} \min\begin{pmatrix} VModeValue(sys), \\ VModeValue(ref_{sys}) \end{pmatrix} \cdot \prod_i W_{Verr-attribute(i)} & \text{when mapped} \\ VModeValue(sys) \cdot W_{V\text{-}FA} & \text{when event not mapped} \end{cases}$$

$$Argument\_Value = Element\_Value(sys\_arg \Rightarrow ref\_arg) \cdot W_{Aerr\text{-}role} + \begin{pmatrix} Element\_Value(sys\_arg \Rightarrow ref\_arg) - \\ Element\_Value(sys\_arg \Rightarrow sys\_arg) \end{pmatrix} \cdot W_{A-FA}$$

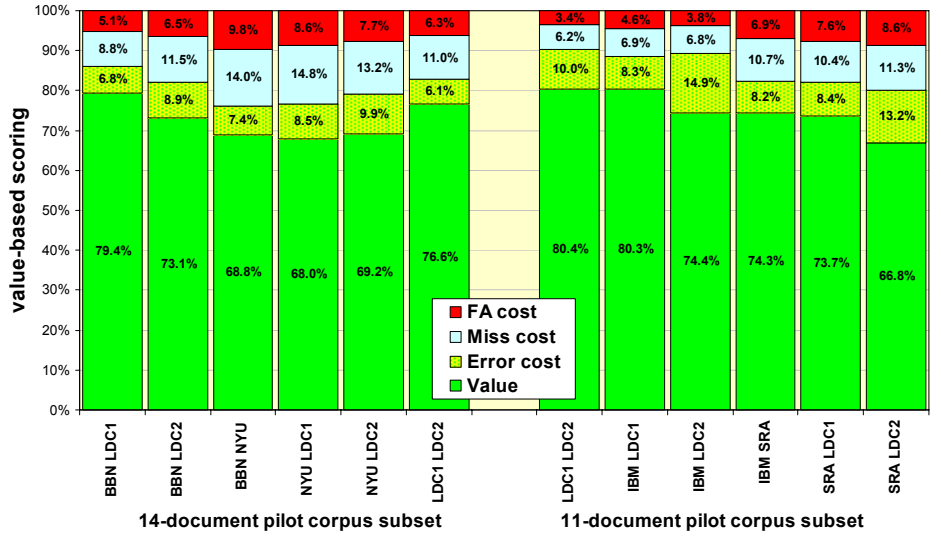
Pilot Corpus Annotation Scores



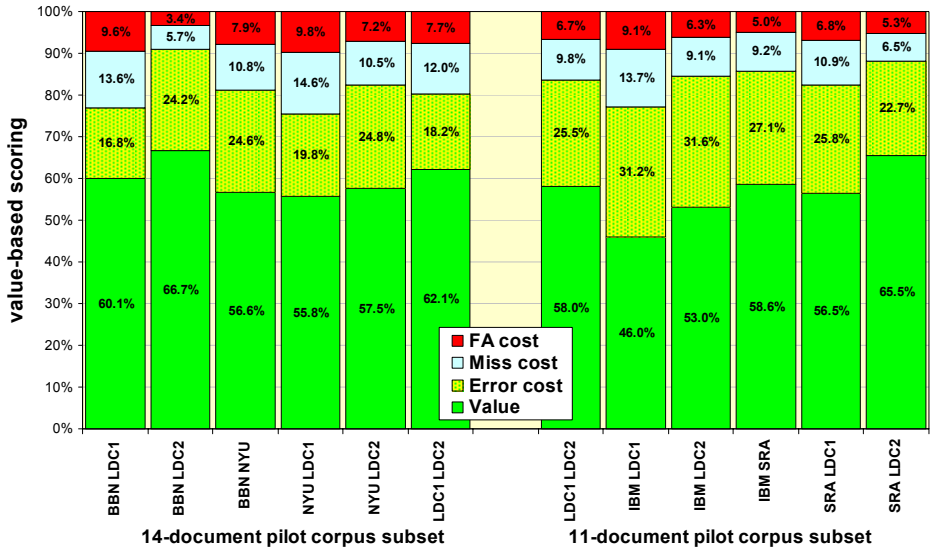
Pilot Corpus Annotation Scores



Average Cross-Site Scores for Relation Annotation



Interannotator Performance on Event Annotation



Interannotator Performance on Quantity Annotation